Journal of Education and Practice
ISSN 2222-1735 (Paper) ISSN 2222-288X (Online)
Vol.16, No.13, 2025

Scaling Course Development with an AI Automation with Human-in-the-Loop System: Contractor Perceptions of Workflow, Scalability, and Quality

Brandi Robinson (Corresponding author)
Department of Educational Leadership, Bowie State University
14000 Jericho Park Rd, Bowie, MD, USA 20715
E-mail: brobinson@bowiestate.edu

Abstract

Recent advances in generative artificial intelligence (AI) have opened new possibilities for accelerating course production in higher education, yet they have also raised concerns about instructional coherence, alignment, and ethical oversight when content is generated at scale. These tensions are especially consequential in high-stakes fields such as nursing, where poorly sequenced or misaligned learning materials can affect clinical readiness and patient safety. As institutions face mounting pressure to expand online offerings quickly, there is a growing need for development models that preserve pedagogical rigor while leveraging automation for efficiency. The rapid growth of artificial intelligence (AI) in higher education has intensified the need to strike a balance between scalability and pedagogical rigor. This convergent mixed-methods study examined how an AI-enabled automation and human-in-the-loop (HITL) system, grounded in the Learner Journey Framework (LJF), supported large-scale course development in a nursing curriculum redevelopment initiative. Quantitative survey data from eight instructional design contractors assessed workflow efficiency, mapping clarity, scaffolding, and perceived rigor, while qualitative interviews explored their lived experiences with the AI-assisted process. Across nineteen courses comprising more than 22,000 instructional assets, participants reported high consensus that the workflow was efficient (Mean [SD] = 4.81 [0.29]), clearly structured (4.78 [0.34]), and pedagogically sound (4.70 [0.33]). Thematic analysis revealed five overarching themes: distributed expertise and structural scaffolding, dialogic human–AI interaction, objective-driven prompting, scalability through structured collaboration, and scaffolding as a learner-pathway mechanism. Together, these findings demonstrate that when automation is disciplined by instructional theory, scalability and rigor can coexist. The LJF's objective-based structure enabled AI agents and human reviewers to co-produce coherent, ethically transparent, and inclusive learning materials at a rapid pace. This study contributes empirical evidence that pedagogically governed AI systems can sustain instructional alignment and ethical oversight at scale, reframing automation as an extension of pedagogy rather than a threat to it.

Keywords: Instructional design, artificial intelligence, human-in-the-loop, scalability, course development, nursing education, Learner Journey Framework, pedagogical alignment, ethics, mixed-methods research

DOI: 10.7176/JEP/16-13-11

Publication date: December 30th 2025

1. Introduction

Institutions are under mounting pressure to scale course production rapidly to meet enrollment growth, cost constraints, accreditation mandates, and workforce demands. Traditional instructional design models are pedagogically rigorous but resource-intensive and therefore struggle to meet institutional demands (Archibald & Feldman, 2016, 2018; Raisman, 2013). This tension between scaling course production and maintaining institutional fidelity is most often felt in high-stakes disciplines, such as nursing, where instructional misalignment directly affects professional readiness and patient safety (Han et al., 2025; McCormick, 2025). Although emerging artificial intelligence (AI)-assisted design systems promise speed through automation, it often comes at the expense of alignment, coherence, and instructional depth (Li et al., 2025; C. Yao et al., 2025).

The challenge for higher education is not whether to use automation, but how to do so without compromising quality and pedagogical integrity. Prior studies demonstrate that generative AI can produce content at scale, emphasizing technical efficiency; however, it often lacks the pedagogical knowledge necessary to ensure alignment between learning outcomes, cognitive progression, and assessment (Schleiss et al., 2023; C. Yao et al., 2025). Systems that rely on content-first generation frequently produce repetitive outputs disconnected from learning objectives or scaffolding sequences. As Li et al. (2025) note, automation within artificial intelligence yields “rapid reproduction of surface structure without depth of learning logic.” This pattern demonstrates the need for frameworks that train AI systems to operate within pedagogical architectures rather than apart from them.

Studies using human-in-the-loop (HITL) AI models in education, where human oversight and machine generation operate collaboratively, have consistently shown that the quality and fairness of courses improve when human judgment is present across the prompting, drafting, and revision stages, rather than being confined to final checks (Memarian & Doleck, 2024). Recent work further argues that many so-called HITL systems are better understood as AI-in-the-loop arrangements, in which humans remain the primary decision makers and AI functions as an efficiency layer, and that evaluation should therefore center human goals and contexts rather than model performance alone (Natarajan et al., 2025). However, research on HITL systems has largely remained conceptual (Mosqueira-Rey et al., 2022; Retzlaff et al., 2024). This leaves a significant gap in evidence about how such systems function within large-scale design workflows (Memarian & Doleck, 2024). In addition, ethical models such as ARCHED (Li et al., 2025) and principles articulated by Nguyen et al. (2023) advocate for transparency and human accountability. However, few studies have examined the practical operationalization in production settings. The absence of empirical validation for pedagogically governed HITL systems thus represents a critical deficiency, particularly in professional education, where scalability must coexist with rigor and ethical oversight.

Therefore, the purpose of this study is to evaluate how a proprietary AI + HITL system, architected using the Learner Journey Framework (LJF), supports the development of scalable and high-quality courses in nursing education. Given that prior research often separates production metrics from practitioner experience, the current study adopted a convergent mixed-methods approach that allows integration of quantitative indicators of scalability with qualitative insights into cognitive and procedural collaboration. Quantitative survey data assessed contractors’ perceptions of workflow efficiency, clarity, and alignment, while qualitative interviews captured their experiences and professional judgments. Each domain was analyzed independently and then merged to identify areas of convergence, where statistical patterns and lived experiences aligned, and divergence, where differing perspectives revealed nuance. This approach aligns with Creswell’s assertion that convergence strengthens validity by enabling complementary insights to confirm or challenge one another (Creswell & Creswell, 2022).

2. Background

2.1 Theoretical Foundations of Learner Journey Framework

The LJF was designed to address the tension between the need to rapidly scale course production while maintaining its pedagogical rigor by encoding instructional theory into the structure of course design (Robinson, 2024, 2025a, 2025b). The five pillars of the framework, Cognitive Sequencing, Embedded Scaffolding, Aligned Assessment, Institutional Consistency, and Durable Learning Pathways, serve as both a pedagogical foundation and production blueprint. When integrated with AI systems, these same principles constrain automation through pedagogical governance, transforming design theory into a quality-assurance mechanism.

LJF is grounded in four key theories: Constructivism (Bruner, 1960), Vygotsky’s Zone of Proximal Development (Vygotsky, 1978), Bloom’s Revised Taxonomy (Anderson & Krathwohl, 2001), and Backward Design (Wiggins & McTighe, 2005). Vygotsky’s concept of guided challenge (Vygotsky, 1978) and Bloom’s taxonomy of cognitive progression (Anderson & Krathwohl, 2001) both emphasize that growth occurs when learners move through intentional sequences of support and release. More recent research by Lungu (2025) demonstrated that backward-aligned course structures improve learner transfer and engagement by making objectives the organizing mechanism of both instruction and assessment. Dazeley *et al.* (2024) extend this principle through *Agile Backward Design*, emphasizing iterative collaboration and stakeholder feedback as essential for scaling design without eroding coherence. Allen (2022) reframes this process as a dialogic exchange between learner and mediating system, suggesting that digital scaffolds—like human guides—can extend cognitive apprenticeship when designed for reflection rather than repetition. Conversely, Roubides and Roubides (2025) warn that constructivism applied indiscriminately in skill-based online contexts may overwhelm rather than support learners, reinforcing the need for structured sequencing as a form of cognitive and ethical care.

Although LJF is rooted in the same constructivist principles that underpin much of modern learning theory, it translates these ideas into a practical design system. The LJF operationalizes these principles by linking objectives, activities, and assessments in a visible progression that helps both instructors and learners manage complexity. Quality, in this view, is not measured by volume or novelty, but by how clearly each element connects to the next in a learner's path toward independence (Robinson, 2024, 2025a, 2025b). Scaffolding functions as the framework's central mechanism. Each objective prepares the ground for the next, and each assessment confirms readiness to continue. When this logic is embedded in course design, coherence becomes a teachable concept. Structure does not constrain creativity; it gives it shape. Similar ideas are also explored in research on human-centered and AI-supported design, where transparency and shared responsibility are identified as key conditions of quality (Li *et al.*, 2025; Memarian & Doleck, 2024). Together, these studies affirm that structure and adaptability can coexist when guided by a coherent learning map.

LJF also extends constructivism through an ethical lens. When objectives and progressions are transparent, rigor becomes predictable rather than arbitrary, which sustains learner persistence and a sense of belonging. In this sense, alignment is not only a technical method but also an ethical stance. Structured design has been associated with greater equity in large-scale and technology-rich environments because it makes expectations explicit and feedback continuous (Cotilla Conceição & van der Stappen, 2025). These relationships between structure, belonging, and transparency mirror the broader call for responsible AI in education, which prioritizes human oversight and interpretability as foundational to trust (Nguyen *et al.*, 2023). Alignment, therefore, becomes a moral as well as pedagogical commitment.

The broader contribution of LJF lies in making its foundational principles visible and repeatable—a shared language for coherence that can scale across programs and disciplines. The empirical applications of the LJF in writing and general education courses demonstrated measurable gains in coherence, student persistence, and design efficiency (Robinson, 2024, 2025a). This structure now serves as the foundation for the present inquiry into AI-supported design. The essential question is no longer whether technology can produce learning materials, but whether such systems can preserve the coherence and care that make learning sustainable and effective. The framework offers that grounding by positioning design as both an intellectual and ethical act. As Schleiss *et al.* (2023) contend, frameworks embedding domain-specific logic are essential to prevent "algorithmic drift" in AI-mediated course production. The LJF operationalizes this principle by linking every instructional asset to a mapped hierarchy of objectives, enabling automation to act as a structured collaborator rather than a generative risk. As a result, the LJF organizes learning as a sequenced progression of both cognitive and applied mastery.

2.2 Ethics, Inclusion, and Human Oversight

In addition to scalability concerns, researchers increasingly emphasize the ethical and inclusive dimensions of AI-assisted design. Nguyen *et al.* (2023) and Bibi (2024) identify transparency, human agency, and fairness as foundations of trustworthy AI integration. Cotilla Conceição and van der Stappen (2025) highlight inclusivity as an ethical requirement while cautioning that unexamined automation risks perpetuate representational bias.

Inclusive learning in AI contexts is facilitated by design choices that provide a visible structure and access to content. Cotilla Conceição and van der Stappen (2025) argue that the systems that promise personalization in higher education practice can also perpetuate exclusion when data provenance and validation are limited. Complementary work on educational AI ethics identifies transparency and human agency as the operational conditions for inclusion, as they position people to identify and correct bias during design rather than after deployment (Nguyen *et al.*, 2023). At the same time, efficiency pressures can mask these equity requirements, which means that speed can exacerbate inequity if the structure is not explicit (Francis *et al.*, 2024). What remains unclear in the literature is how inclusion is enacted inside day-to-day course design workflows rather than at the policy level. LJF addresses this gap by transforming equity into a structural practice, mapping objectives and scaffolds to ensure access and progression are designed in from the outset (Robinson, 2024, 2025a, 2025b).

Francis *et al.* (2024) extend the arguments on inclusivity as an ethical requirement to policy, urging institutions to strike a balance between innovation and integrity through sustained human verification. In inclusive AI, human oversight functions as the mechanism that converts ethical intent into design behavior. Mixed-methods evidence shows that educators and students describe trustworthy AI as transparent and reviewable, with clear records of what the system produced and what a human approved (Bibi, 2024). Reviews of HITL models in education consistently find improvements in quality and fairness when human judgment guides prompting, drafting, and revision, rather than just being involved during the final review (Memarian & Doleck, 2024).

Frameworks that formalize shared accountability, such as ARCHED, add that collaboration and explainability must be encoded as process requirements, not preferences, if equity goals are to persist under production pressure (Li *et al.*, 2025). Even so, few studies demonstrate how such oversight reshapes the course-level experience for learners in high-stakes disciplines. Within this discourse, the LJF serves as both a pedagogical and an ethical infrastructure, embedding inclusivity and representation directly into the design logic. By aligning every asset to multi-tier objectives, the LJF provides observable checkpoints where human review can verify clarity, relevance, and difficulty for diverse learners, thereby operationalizing inclusion within the workflow rather than around it (Robinson, 2024, 2025b).

The equity question is most acute where both accuracy and pace are crucial, such as in the health sciences, including nursing. Program reports describe curriculum initiatives that combine concept-based frameworks with AI-supported authoring to manage scope without compromising rigor, noting perceived gains in coherence and student preparedness when structure drives production (McCormick, 2025). Studies of learner experience suggest that the acceptance of AI-supported materials increases when tasks are scaffolded and expectations are explicitly stated, which ties inclusion to transparency and cognitive progression in concrete ways (Han *et al.*, 2025).

Planning models for AI course design also demonstrate that when sequencing and alignment are specified upfront, teams can scale their output with improved usability and consistency, suggesting that structure performs the ethical work by maintaining purpose visibility during generation (Schleiss *et al.*, 2023). Still missing is empirical evidence that links these structural moves to the perceptions of quality and to the overall coherence of courses when AI drafts are reviewed at scale. The present study addresses that missing layer by examining how contractors, working within the LJF and a HITL system, perceive scalability alongside rigor, alignment, and inclusivity in complete nursing courses (Robinson, 2024, 2025a).

Across these lines of inquiry, inclusion is not separate from quality; it is quality made visible. The scholarship converges on transparency, human oversight, and structural alignment as the conditions that sustain equity in AI-enabled design (Bibi, 2024; Nguyen *et al.*, 2023). Frameworks that encode those conditions into everyday workflow provide the most durable path to ethical scale because they translate principles into practice under production constraints (Li *et al.*, 2025; Schleiss *et al.*, 2023). What remains to be shown is how a single, theory-aligned system performs when teams use it to build multiple courses quickly and then evaluate finished courses for coherence and rigor. The LJF offers that test case because it treats alignment and scaffolding as equity practices, making inclusion an inherent property of design rather than an afterthought (Robinson, 2024, 2025b).

2.3 Human–AI Collaboration and Course Design Quality

The integration of AI into instructional design continues to transform both the pace and scope of educational development. AI systems now produce draft materials with unprecedented speed, yet efficiency alone cannot ensure coherence or theoretical integrity. Much of the early experimentation with generative tools has centered on *content-first generation*—outputs that appear complete but lack pedagogical sequencing or objective alignment. Such materials may accelerate production but often fragment the learner experience, producing courses that feel assembled rather than designed (Robinson, 2025a; C. Yao *et al.*, 2025). The persistent challenge is not the technology's capacity but its capacity for *theory-informed intention*. As Schleiss *et al.* (2023) observe, scalable course design succeeds only when automation is bounded by a framework that defines learning purpose, cognitive progression, and assessment fidelity. Learning still depends on human interpretation and the structural logic that links tasks to mastery.

HITL models respond directly to this tension. In such systems, human expertise is integrated throughout every phase of automation, rather than being added as a post-hoc review. Designers guide the AI's prompts, evaluate its drafts, and refine outputs to sustain conceptual coherence. This iterative rhythm parallels the cycles described in Agile Backward Design, where development proceeds through repeated plan–develop–evaluate loops that maintain alignment across rapid production cycles (Dazeley *et al.*, 2024). Embedding such agile logic within human–AI workflows operationalizes what Lungu (2025) terms "design as discipline," ensuring that automation accelerates precision rather than replacing reflection. Recent studies emphasize that these collaborations perform best when guided by transparent frameworks that articulate both pedagogical intent and quality benchmarks (Li *et al.*, 2025; Memarian & Doleck, 2024). Within this space, the LJF functions as the necessary scaffold. It supplies the AI with a clear map of objectives, cognitive levels, and alignment patterns—essentially a design grammar that automation can follow. In doing so, the framework positions humans and machines as complementary co-designers: the AI extends reach, while the human preserves meaning.

Quality in these hybrid environments must therefore be defined as both *process* and *product*. Coherence across objectives, activities, and assessments remains the most accurate indicator of rigor, while scaffolding regulates the pace and difficulty of learning. Empirical work in medical and nursing education demonstrates that when structured frameworks guide AI-supported development, outcomes improve in both clarity and learner engagement (Han *et al.*, 2025; McCormick, 2025). H. Yao *et al.* (2025) further report that "full co-pilot" modes—continuous human oversight during generation—yield the highest instructional quality scores, echoing the LJF's principle that alignment must remain visible and interactive throughout the design process. Together, these studies affirm that automation can scale design only when paired with intentional structure.

At the same time, collaboration between humans and AI raises questions of transparency and accountability. Frameworks such as ARCHED stress that ethical and human-centered design depends on interpretable systems and traceable decisions (Li *et al.*, 2025). Similar positions are also presented in the broader ethics literature, where scholars argue that responsible AI should prioritize institutional trust over efficiency alone (Bibi, 2024; Francis *et al.*, 2024; Nguyen *et al.*, 2023). When the LJF governs the workflow, these concerns become built into the process itself. The framework's mapped objectives and feedback loops ensure that every AI-assisted artifact can be traced to a pedagogical intention. What results is not merely a faster workflow but one that is ethically legible—where alignment doubles as assurance.

In this study, such intersections of automation, structure, and human oversight form the central point of inquiry. The LJF provides a theoretical and procedural framework for assessing how contractors perceive scalability, quality, and rigor within an AI-supported system. It also grounds the study's attention to experience—how designers interpret the balance between autonomy and guidance when working with automation. By centering both the structural and ethical conditions of collaboration, the research extends the conversation on AI in education beyond questions of efficiency toward those of fidelity, transparency, and trust. Ultimately, the framework positions design as a shared act of precision and care—one that technology may accelerate but never replace.

2.4 Integrative Synthesis and Conceptual Gap

Constructivist theory explains how learning develops through guided challenge and purposeful scaffolding (Vygotsky, 1978). The LJF translates these principles into a visible structure that links objectives, activities, and assessments through deliberate progression. Artificial intelligence and HITL collaboration expand what can be created at scale, yet they also risk weakening coherence when design is driven by automation rather than theory. Ethical models of transparency and inclusion help sustain trust and a sense of belonging in these evolving systems. What remains uncertain is how structured, theory-aligned frameworks, such as the LJF, perform within AI-supported design environments. This distinction matters because sustainable innovation depends on alignment that preserves both rigor and care.

While scholars have developed models for multi-agent AI collaboration (Li *et al.*, 2025; C. Yao *et al.*, 2025), few studies have analyzed how human contributors and automation interact to maintain instructional fidelity during rapid production. Memarian and Doleck (2024) call for "mutual calibration" between human and machine reasoning—an iterative, pedagogically grounded process of alignment—but empirical demonstrations of this mechanism remain scarce. This study directly addresses that gap by examining how a theory-trained multi-agent AI system, integrated with structured human oversight, functioned across nineteen nursing courses. In this study, the AI + HITL workflow positions human verification as an embedded ethical function rather than a final checkpoint where contractors review AI-generated assets for tone, inclusivity, and disciplinary accuracy, ensuring alignment with both instructional intent and professional representation. Such structural oversight aligns with the definition of ethical integration proposed by Nguyen *et al.* (2023), which emphasizes the integration of accountability and transparency into the process, rather than merely appending them as a compliance measure.

3 Methodology

3.1 Research Design

This study was guided by the following three research questions to examine how an AI automation and HITL workflow supported scalable, high-quality course development using the LJF:

1. How do contractors perceive the role of an AI automation + HITL system in developing nursing courses at scale?
2. What are contractors' experiences working with the automation system and reviewing AI outputs through the HITL process?
3. How do contractors evaluate the scalability, quality, and alignment of courses produced through this system compared to traditional development models?

These questions collectively examine both the process (workflow experience) and product (course quality and alignment) dimensions of AI-enabled instructional design, aligning directly with the study's purpose to evaluate scalability, quality, and professional judgment within the LJF.

To address these questions, a convergent parallel mixed-methods design, as proposed by Creswell and Creswell (2022), was adopted. The quantitative component consisted of survey items assessing contractors' perceptions of workflow scalability, clarity, scaffolding, rigor, and alignment. The qualitative component included open-ended survey responses and semi-structured interviews exploring contractors' experiences and professional judgments. This design captured both the measurable outcomes of workflow efficiency and the contextual insights that explain *why* and *how* those outcomes emerged.

Grounded in a pragmatic mixed-methods paradigm (Creswell & Creswell, 2022), this design emphasized the integration of empirical measurement to illuminate breadth and qualitative narratives deepened interpretation, ensuring that evidence of scalability was paired with the human meaning-making behind it. Integration of both strands strengthened validity through triangulation and complementarity. Pragmatism supports the study's applied goal—to evaluate both efficiency and pedagogical coherence in a real-world instructional-design context where practical outcomes are as critical as theoretical contribution.

3.2 AI Model

A proprietary, multi-agent AI-assisted course-development system, trained on the LJF, was used to generate first-draft instructional assets that were refined by human reviewers to ensure pedagogical alignment, accuracy, and rigor. The model training data and the human quality-assurance process were guided by the five pillars of LJF, namely cognitive sequencing, embedded scaffolding, aligned assessment, institutional consistency, and durable learning pathways (Robinson, 2024). Thus, the framework functioned as both an architectural blueprint and a conceptual lens, translating learning theory into an operational design logic that governed automation, human review, and final course assembly.

The proprietary system comprised approximately thirty specialized AI agents, each designed to execute a distinct instructional-design function—for example, generating text pages, aligning assessments, creating visuals, or structuring formative feedback. Each agent's prompting logic and evaluation criteria were derived from the LJF's interpretation of that task, ensuring that automation operated within the framework's constructivist learning architecture rather than apart from it. This multi-agent orchestration reflected the LJF's principle of distributed cognition, allowing automation to function as a coordinated pedagogical ecosystem rather than a single generative model.

Throughout production, the researcher conducted iterative pedagogical and technical tuning of prompts and model behaviors to preserve fidelity between automation and instructional intent. This dual calibration process paralleled the system's own feedback loops—human corrections informed prompt refinement, while updated agent behaviors improved subsequent AI outputs—embodying the continuous alignment described by Memarian and Doleck (2024) as essential to trustworthy HITL design.

Approximately 22,800 instructional assets were produced across 19 nursing courses (~ 1,200 per course). Each course followed a ten-module structure, and every module contained an average of 60–80 integrated assets, including text pages, interactives, assessments, videos, case studies, and discussion activities. Assets were explicitly mapped to module-level and enabling-level objectives within a three-tier system (program competency → module objective → enabling objective), ensuring vertical and horizontal alignment between learning outcomes, activities, and assessments.

3.3 Participants

The study took place during a 22-week nursing curriculum redevelopment project at an accredited U.S. university. Eight instructional design contractors directly involved in the nursing curriculum redevelopment project participated in the study. Participants met two inclusion criteria: (a) direct involvement in all 19 course builds employing the AI + HITL workflow and (b) completion or review of at least one instructional-asset type within each course (e.g., text pages, assessments, interactives, or graphics). All participants were experienced instructional designers (3–15 years of practice) and provided informed consent prior to participation.

3.4 Quantitative Instrument

A researcher-designed online survey measured six composite constructs: (a) workflow experience, (b) mapping system (module-level and enabling objectives), (c) clarity and alignment, (d) scaffolding and learner support, (e) quality and rigor, and (f) comparative evaluation. Each construct included 4–6 Likert-type items (1 = *Strongly Disagree* to 5 = *Strongly Agree*). Internal-consistency reliability was high across all scales (Cronbach's α = .86–.94).

Each survey included an embedded module-review component, in which participants examined three completed modules containing all instructional assets—text pages, interactives, assessments, and discussion activities—to ground their quantitative ratings in a direct evaluation of the full course's coherence and instructional flow. The complete instrument is presented in **Table 1**.

3.5 Qualitative Instruments

Qualitative data were gathered through open-ended survey items and semi-structured interviews guided by five domains: workflow experience, scalability, mapping and alignment, module clarity and scaffolding, and reflections on professional judgment. Interviews, lasting 45–60 minutes, were conducted via secure Zoom, recorded, and transcribed verbatim for analysis.

3.6 Data Collection

As part of the survey, each contractor reviewed three fully completed modules, which contained all instructional assets, including text pages, interactives, assessments, and discussion activities. This process ensured that participants could evaluate the coherence and quality of the entire instructional sequence, rather than in isolation. Viewing the assembled modules allowed them to see how the "assembly line" of AI-generated and human-refined assets came together to produce a finished learning experience. Their survey responses and subsequent interviews, therefore, reflected judgments informed by direct exposure to complete, aligned course structures.

Interviews were conducted through Zoom, recorded locally, and transferred immediately to encrypted Box storage. Original recordings were deleted from Zoom once transcription and verification were complete. All transcripts were stored and coded in Dedoose, an encrypted qualitative analysis platform.

Data collection spanned four weeks during the final phase of course production. In this way, the timing of data gathering coincided with participants' active engagement in course design, ensuring the authenticity of reflection and the immediacy of the experience. Quantitative and qualitative data were collected concurrently, analyzed separately, and merged during interpretation to provide a comprehensive understanding of both process efficiency and product quality.

3.7 Quantitative Analysis

Descriptive statistics (means, standard deviations, and ranges) were computed for each composite construct using Likert-scale survey data. Composite scores were calculated by averaging items within each construct—Workflow Experience, Mapping Clarity and Alignment, Scaffolding and Learner Support, Perceived Rigor and Quality, and Comparative Evaluation. Higher mean values indicated stronger agreement with statements about efficiency, alignment, and instructional quality.

Consistent with the guidance for exploratory mixed-methods analysis (Creswell & Creswell, 2022), quantitative results were first examined descriptively to establish general patterns that could later be compared with qualitative insights. This phase emphasized trend identification and internal reliability rather than hypothesis testing, reflecting the study's pragmatic aim of understanding how contractors perceived workflow efficiency and course quality. Quantitative findings addressed all three research questions.

Survey data were analyzed descriptively and then compared with qualitative narratives to identify convergence between numerical trends and participants' reported experiences. Reports of limited revision needs and consistent alignment, gathered through interviews, were used to corroborate patterns observed in quantitative scores.

3.8 Qualitative Analysis

Open-ended and interview data were analyzed using thematic analysis (Braun & Clarke, 2008). The process included: (a) familiarization with transcripts, (b) open and theory-informed coding aligned with LJF pillars, (c) collapsing codes into categories, (d) refining and naming themes, and (e) validating coherence across data.

Three trained externs independently coded transcripts in Dedoose®, following the triangulation model used in previous LJF studies (Robinson, 2025b). Discrepancies were resolved through consensus meetings and peer debriefing. Coding and theme development were accompanied by memo writing that tracked the interpretive evolution and maintained an audit trail of analytic shifts, consistent with the notion that qualitative rigor arises from reflexivity, triangulation, and transparency of analytic decisions (Creswell & Creswell, 2022).

3.9 Integration of Data

Following an independent analysis of both quantitative and qualitative strands, the findings were integrated through joint display triangulation and meta-inference (Creswell & Creswell, 2022). Quantitative composites (e.g., Workflow Effectiveness, Mapping Clarity, and Perceived Quality) were compared with corresponding qualitative themes and illustrative quotes.

Convergence occurred where qualitative descriptions of clarity, structure, and efficiency reinforced high quantitative ratings. For example, strong mean scores for workflow effectiveness aligned with repeated statements that the process was "organized without losing clarity." Divergences—instances where participants praised speed but discussed early tuning challenges—were examined to explain nuance and developmental patterns within the workflow.

This integration followed Creswell's emphasis on synthesis as the hallmark of mixed-methods validity, where meaning emerges through the relationship between strands, rather than their isolation (Creswell & Creswell, 2022). The joint interpretation thus functioned as both validation and sense-making, illustrating how AI-assisted course design achieved scalability and rigor simultaneously when grounded in a structured pedagogical framework.

3.10 Data Security and Ethical Considerations

The Bowie State University IRB approved this study (IRB # 491-304). Data security procedures adhered to institutional and federal ethics standards. Participant consent forms were collected via encrypted Box Drive; recordings were stored securely and deleted from Zoom once verified; and transcripts and codes were housed in Dedoose, which provides AES-256 encryption and multi-factor authentication.

Participation was voluntary and uncompensated. Survey data were collected anonymously, and interview transcripts were de-identified to ensure confidentiality. Findings are presented in aggregate or through anonymized quotations only. No deception or physical procedures were used, and participant risk was minimal.

3.11 Researcher Role and Trustworthiness

The researcher served in a dual role as the curriculum architect and framework developer for the LJF, but did not participate in project execution or supervision of the contractor. This provided contextual depth while requiring transparency regarding analytic distance—a balance that Creswell describes as the *informed external–internal stance* (Creswell & Creswell, 2022).

Potential bias arising from the researcher's role in framework development was mitigated through multiple strategies. First, reflexive memoing documented analytic decisions and interpretive shifts throughout data analysis. Second, reflexive journaling captured the researcher's positional reflections, assumptions, and emotional responses as they navigated the dual perspective of framework creator and independent analyst. Third, external coder verification and peer debriefing were employed to assess the stability of emerging themes and mitigate interpretive drift.

Procedural and interpretive validity (Creswell & Creswell, 2022) were strengthened through triangulation, peer review, and the maintenance of an encrypted audit trail that linked memos, coding iterations, and analytic decisions. Because participants were independent contractors, formal member checking was not feasible; however, cross-validation through an external coder review functioned as an equivalent method for confirming the analytic credibility and coherence of themes.

These practices together formed a comprehensive reflexivity system that unified memoing, journaling, peer review, and audit documentation. This integration ensured both procedural rigor (what was done) and interpretive integrity (how meaning was constructed). In Creswell's framework, such transparency and reflexive engagement constitute hallmarks of qualitative trustworthiness within mixed-methods research (Creswell & Creswell, 2022).

This reflexive stance also extended to ethical verification, ensuring that analytic processes mirrored the same HITL principles guiding the study's design. Oversight, transparency, and continuous reflection were not merely analytic practices—they embodied the study's central ethos: that both AI systems and researchers achieve integrity through structured human judgment.

4 Results

Quantitative results reflected a strong consensus that the workflow was efficient, structured, and pedagogically clear. Process analytics confirmed that the AI + HITL workflow accelerated production by approximately 234% compared to historical course-development baselines, while sustaining correction rates below 5%. The increase reflected structural efficiency rather than compression of design quality—evidence that disciplined automation can multiply throughput without eroding pedagogical integrity. Workflow Effectiveness achieved the highest mean (M = 4.81, SD = 0.29), followed by Mapping Clarity and Alignment (M = 4.78, SD = 0.34), Scaffolding and & Learning Support (M = 4.88, SD = 0.35) and Perceived Rigor and Quality (M = 4.70, SD = 0.33). Low variance across constructs (SD < 0.36) suggested consistent experiences of predictability and control. Similarly, participants described the workflow as both structured and adaptive—a production environment that accelerated creation without compromising instructional rigor, as noted in qualitative interviews.

Overall, the following four themes emerged as a result of the integration of quantitative and qualitative data: (a) workflow efficiency and human oversight, (b) mapping as the framework's core engine, (c) clarity, scaffolding, and learner progression, and (d) quality, consistency, and confidence in AI outputs.

4.1 Theme 1 – Distributed Expertise and Structural Scaffolding

Participants consistently emphasized that dividing production into clearly defined roles enhanced both speed and precision. "Because I only had to focus on one piece—videos—I could make them stronger," noted one contributor (P2). Others described the process as "an assembly line that still felt creative" (P6). The same participant also noted that "It stayed organized without losing clarity" (P6). Another contractor emphasized that "the whole process was smooth and consistent" (P8). These perceptions corresponded with the highest workflow-efficiency rating (M = 4.81, SD = 0.29) and reinforced the core LJF principle that clarity of cognitive contribution is foundational to quality at scale (Robinson, 2025b). Internal revision logs corroborated these impressions, showing fewer than five percent substantive edits across all outputs—evidence that accuracy and alignment were sustained even under rapid production schedules.

Rather than fragmenting learning design, the distributed model acted as a form of cognitive scaffolding. Each specialist advanced a discrete step in the instructional sequence while maintaining awareness of the whole. Open-ended survey responses echoed this structure—describing the process as "smooth," "predictable," and "organized without losing clarity." Low variance across workflow and alignment items (SD < 0.35) indicated a shared experience of reliability. This pattern parallels Li *et al.* (2025), who observed that structured multi-agent collaboration stabilizes quality in AI-assisted instructional design.

Participants also reported unexpected pride in collective craftsmanship. "It was nice to see it all together in Blackboard—it looked cohesive," one participant reflected (P7). Another participant added that "Even though we each did a small part, it felt like one voice" (P8). As one respondent summarized, "It wasn't just faster—it made us better" (P1). This affective ownership challenges assumptions that assembly-line processes diminish professional identity. It aligns with Francis *et al.* (2024) and Cotilla Conceição and van der Stappen (2025), who found that inclusive, ethically collaborative design practices heighten engagement and accountability. Together, these accounts demonstrate that efficiency emerged not despite, but through, craftsmanship. The LJF's objective-driven segmentation transformed what might have been mechanical division into distributed cognition, demonstrating that when instructional logic structures production, scalability, and pride can coexist.

4.2 Theme 2 – Dialogic Human–AI Interaction

Participants described the interaction between the AI system and human contributors as an evolving partnership that balanced automation with active guidance from the architect of the LJF and the multi-agent AI system. Early in the project, several noted the system's responsiveness: "It was almost like having a super SME that's right there with you all the time" (P6). The AI served as a constant collaborator—available instantly for clarification and revision—while the HITL step preserved accuracy and alignment. "I would not only be able to ask the question in real time; I'd get an answer in real time. I didn't have to wait," explained one participant (P6). Compared with traditional builds, "in previous projects, it would take me three months to do a single course… here I could build multiple courses… and then a week later, I have all of her, 'yep, looks good'" (P6).

Other contributors described a learning curve that transformed into confidence as the workflow matured. As one participant reflected, "You have to trust the system… the more you trust it, the better the process is" (P7). Another participant recalled, "It started to learn exactly what I was going to ask for" (P8), noting an improvement in prompt responsiveness over time. Participants attributed these gains to continuous tuning—both pedagogical and technical—led by the LJF's architect: "[She] would respond in seconds—'go in this one, try this'—and then it kept going smoothly" (P8). In practice, this meant that tuning occurred not only through end-user feedback but also through continuous architectural adjustment of agent behavior and prompt libraries to sustain fidelity with pedagogical intent.

These accounts illustrate what Memarian and Doleck (2024) describe as mutual calibration in HITL systems, where quality emerges from iterative alignment between human oversight and machine responsiveness. The pattern also parallels Li *et al.* (2025) and Edirisingha (2024), who found that reliability and efficiency improve when iterative cycles are transparent and pedagogically anchored. Workflow ratings remained high (M = 4.81, SD = 0.29), and revision logs showed substantive corrections averaging below 5%—evidence that iterative tuning enhanced accuracy without eroding speed. Participants emphasized that agility did not replace human review: "making sure it said patient when it needed to, not learner," (P6). This aligns with Nguyen *et al.* (2023) and Bibi (2024), who position human judgment as the safeguard that transforms efficiency into integrity. The workflow thus functioned as adaptive collaboration—a "super SME on call"—compressing production cycles while maintaining pedagogical precision and ethical transparency.

4.3 Theme 3 – Objective-Driven Prompting and Pedagogical Alignment

Mapping Clarity and Alignment achieved one of the highest quantitative composites (M = 4.78, SD = 0.34). Participants consistently identified the hierarchy of module-level and enabling objectives as the cornerstone of accuracy and efficiency. "When I added the enabling objectives, the AI output finally made sense," one participant (P5) explained. Others noted that "having that mapping built in kept the AI on task—it couldn't wander" (P5). Designers described this structure as "the blueprint" that transformed prompting from creative guessing into predictable instructional generation. Internal logs confirmed the perception: correction rates averaged below 5 percent, and most edits addressed tone rather than content accuracy, demonstrating low variance in quality across modules.

This pattern substantiates Li *et al.* (2025), who found that transparent cognitive scaffolding within human-AI systems increases generative fidelity. Within the LJF, each agent's logic was pedagogically—not merely technically—defined; the framework itself constrained what counted as a "correct" output. As Memarian and Doleck (2024) argue, HITL reliability arises when humans design the evaluative grammar that the AI must follow. Here, that grammar was the objective map. In effect, pedagogical alignment became the quality-assurance layer, allowing the system to generate roughly 1,200 assets per course with measurable precision (SD < 0.35). The constructivist orientation of the LJF—knowledge emerging through structured interaction—ensured that automation amplified, rather than replaced, instructional reasoning. The integration of instructional theory and automation exemplifies Robinson's assertion that learning frameworks can function as operating systems for ethical AI design—each objective serving simultaneously as a learning target and a safeguard against drift (Robinson, 2025b).

4.4 Theme 4 – Scalability through Structured Collaboration and Consistency

Despite the assembly-line workflow, participants described the process as cohesive and motivating. As one participant recalled, "It took three months before to build a course—now I could do multiple in a few weeks," (P6). Others credited shared prompt libraries and tone guides for ensuring visible consistency: One participant noted, "You couldn't tell who wrote which asset," (P3), while another added, "Even though we each did a small part, it felt like one voice" (P8). Workflow Effectiveness again achieved the highest mean, confirming perceptions of speed without loss of clarity.

The sense of team identity and pride in precision challenged assumptions that automation fragments authorship. Contractors framed their contributions as essential links in a pedagogical chain, expressing satisfaction that the system "used everyone's strengths" (P2). This collaborative rhythm illustrates the LJF principle of distributed cognition, where expertise resides in the interaction between humans and tools rather than in individuals. Approximately thirty specialized agents executed discrete, pedagogically defined tasks—from text generation to alignment checks—enabling human reviewers to verify quality in parallel.

This orchestration exemplifies what McCormick (2025) terms concept-based scalability, where clarity of roles sustains consistency across large-scale builds. Comparable dynamics are also observed in Edirisingha (2024), who identifies iterative dialogue as central to the sustainable adoption of AI. The affective cohesion observed here reinforces the LJF's social-constructivist principle that collaborative authorship enhances cognitive presence and motivation. One participant summarized this as follows: "Before, every person did everything, and nothing matched. Now it just fits" (P4). Constructivist alignment again underpinned the success: participants' shared understanding of what "good" looked like—the LJF's alignment literacy—ensured that consistency became a collective cognitive habit rather than an imposed constraint. The structural clarity that aligned creators also shaped the learner experience, producing coherence recognizable across all nineteen courses.

4.5 Theme 5 – Scaffolding and Learner Pathways

Designers portrayed the finished modules as coherent learning journeys that guided novices step-by-step: "It built like a staircase—each part set up the next" (P3). Another participant added, "The images were a great backdrop to reinforce what was in the module—everything aligned for consistency and flow" (P8). Scaffolding and Learner Support averaged 4.88 (SD = 0.35), and open-ended survey responses echoed these ratings, frequently using terms such as "clear," "logical," and "not overwhelming" to describe the module progression. Participants intentionally wrote for the novice learner: "My images are for a novice learner—I never want them to feel overwhelmed" (P8).

Structured sequencing—from text pages to visuals to interactive cases—allowed learners to test comprehension incrementally while maintaining motivation. This deliberate pacing operationalized the LJF's construct of progressive scaffolding, aligning with Edirisingha (2024) on sequenced interactivity as a driver of retention and Li *et al.* (2025) on scaffolded practice as a mechanism for cognitive transfer. The consistency of multimodal reinforcement also reflected the inclusivity described by Cotilla Conceição and van der Stappen (2025): visual and textual parity ensured that diverse learners could engage meaningfully. Participants noted that rigor was preserved through clarity rather than complexity—"It was challenging but achievable—the kind of rigor students respect" (P7). This combination aligns with Francis *et al.* (2024) in striking a balance between innovation and academic integrity.

Each ten-module course contained about 1,200 integrated assets—text pages, videos, interactives, case studies, and Walk-With-Me coaching sequences—produced through human-AI orchestration. Within those assets, scaffolding functioned simultaneously as a design method and learner signal: the rhythm of visuals, text, and dialogue modeled expert reasoning for novices. This echoes the description of scaffolding by Robinson (2025b) as the structural heart of the LJF and its claim that coherence in form creates confidence in learners.

Together, these five themes reveal how the LJF operates as both a pedagogical blueprint and a production protocol, transforming multi-agent automation into a reproducible expression of instructional design theory.

4.6 Cross-Cutting Dimensions: Structure, Cognition, and Ethics

4.6.1 Structural Alignment

Across all strands, participants credited the LJF for converting complexity into predictability. The framework's hierarchical architecture—course objectives decomposed into enabling outcomes and mapped to assets—functioned as what Li *et al.* (2025) describe as a *pedagogical control system*, aligning human and machine logic through explicit design grammar. "It just kept everything tight," one contributor explained, "the AI couldn't drift because the objectives boxed it in" (P5). Quantitatively, this structural consistency appeared in the nearly identical means (~4.8) for Workflow Effectiveness and Mapping Clarity, with dispersion below 0.35 SD—demonstrating uniform confidence that alignment mechanisms held firm even as production scaled to 19 courses and roughly 1,200 integrated assets each.

Participants perceived structure not as a constraint but as creative scaffolding. "Speed only worked because there was structure—AI alone can't do that" (P4). This finding aligns with Schleiss *et al.* (2023), who argue that domain-specific frameworks enable generative systems to operate *within* curricular boundaries rather than outside them. Within the LJF, architecture became the invisible discipline that made creativity safe, ensuring that every agent, asset, and editor operated from a shared pedagogical blueprint. These findings resonate with the argument of Dazeley *et al.* (2024) that agile backward design transforms curricular frameworks into living systems that evolve through stakeholder feedback, and with the evidence from Lungu (2025) that structured alignment fosters equitable access by making progressions visible. Together, they reinforce that scalability grounded in backward and agile design principles constitutes both a pedagogical and an ethical infrastructure.

4.6.2 Collaborative Cognition

The production ecosystem evolved into a distributed reasoning network in which human and machine cognition intertwined. Contractors described the process as "thinking together," a phrase resonant with the conception of *mutual calibration* in HITL systems by Memarian and Doleck (2024). "It was like having a second brain that worked faster but still needed me to make it make sense," one designer reflected (P7). The AI handled pattern recognition and draft assembly, while humans provided interpretive oversight, tone adjustment, and contextual awareness.

Survey items on collaboration and trust correlated strongly with efficiency ($r = 0.83$), indicating that participants perceived cognitive partnership as integral to speed. One participant captured this co-evolution as follows: "It wasn't static—the AI got better as we got better" (P1). Such reciprocal learning illustrates Robinson's contention that constructivist alignment operates not only for learners but also for designers—that shared frameworks transform workflow into a site of professional cognition (Robinson, 2025b). The dialogic exchange of prompting, reviewing, and refining represented an epistemic apprenticeship between human and machine, where understanding emerged through structured dialogue rather than solitary expertise.

4.6.3 Ethical Transparency and Human Verification

Ethical assurance was embedded in the workflow rather than appended at the end. Every asset passed through human verification for tone, representation, and disciplinary accuracy. As observed by one participant, "I always checked—does this sound like a nurse talking to a patient, or like a robot describing care?" (P6). Such vigilance aligns with Nguyen *et al.* (2023) and Bibi (2024), who define ethical AI integration as *continuous human mediation* rather than procedural oversight.

The LJF's objective-driven discipline thus functioned as both an instructional and moral safeguard. Because each output was anchored to a transparent learning target, decisions about inclusion, terminology, and imagery were traceable. "Anyone could see themselves in the scenario," noted one participant (P8), emphasizing representational accuracy and inclusivity. This aligns with the connection between ethical AI design, belonging, and learner identity described by Cotilla Conceição and van der Stappen (2025). Quantitatively, the perception of ethical clarity remained high (M = 4.8, SD = 0.24, paralleling Workflow Effectiveness (M = 4.81), indicating that participants viewed ethical oversight and operational performance as inseparable.

Human verification thus served as what Li *et al.* (2025) term a *responsibility anchor*—a procedural mechanism ensuring that speed never substituted for scrutiny. The combination of automated precision and human interpretation preserved both accuracy and empathy, reflecting Francis *et al.* (2024)'s call for integrity as a co-equal design variable in generative education.

5 Discussion

The findings of this study demonstrate that scalable, ethical, and rigorous course design using AI is achievable when automation is structured through pedagogical governance. The LJF, functioning as both an instructional model and an operational policy, translated theory into a design logic that is enforceable. Its objective hierarchy, defined HITL checkpoints, and traceable production workflows represent not merely tools for efficiency but a model for institutional accountability in generative education.

Each of the nineteen completed courses followed a ten-module structure and contained approximately 1,200 integrated assets, including text pages, interactives, videos, case studies, and Walk-With-Me clinical coaching activities. Contractors consistently attributed their ability to maintain rigor while scaling production to the structural clarity imposed by the LJF. Quantitative convergence and qualitative testimony together substantiate the assertions of Li *et al.* (2025) and Memarian and Doleck (2024), who argue that objective-driven, human-centered frameworks convert automation from a technical shortcut into a *pedagogical amplifier*. As Edirisingha (2024) and Robinson (2025b) similarly contend, alignment, scaffolding, and reflective sequencing are not by-products of quality—they are its preconditions.

Taken together, these dimensions illustrate how the LJF transformed multi-agent automation into a *pedagogically disciplined ecosystem.* Structural alignment provided the foundation, collaborative cognition supplied the mechanism, and ethical transparency anchored oversight. Efficiency, accuracy, and integrity operated as co-dependent outcomes across nineteen courses and approximately 1,200 integrated assets of a single design logic—automation trained by pedagogy and sustained by human judgment. As Edirisingha (2024) observes, generative systems achieve educational legitimacy only when "the design teaches the designer." In this study, that reciprocity was realized: the system learned from its humans, the humans learned from their system, and the framework ensured both learned responsibly.

Our findings have several implications for policy and practice, which are separately discussed below:

5.1 Pedagogical Infrastructure as Policy

Within higher education, policy has traditionally focused on what is taught or who teaches, but rarely on how the *design logic itself governs automation.* The LJF's performance—nineteen courses, approximately 1,200 assets per course, and less than 5 percent average correction rate—suggests that frameworks can serve as compliance mechanisms that regulate AI through instructional intent rather than technical restriction.

This echoes Li *et al.* (2025) and Memarian and Doleck (2024), who argue that human-centered frameworks must encode oversight into system architecture. In the LJF, oversight was not a review layer but a design constant; alignment between objectives and outputs made every artifact auditable. Institutions seeking to operationalize AI ethically should thus treat pedagogical structure as a *policy instrument*—embedding learning frameworks within AI infrastructure to guarantee fidelity, quality, and traceability.

5.2 Ethical Accountability and Human Oversight

The study also reinforces that HITL is not a safeguard added after automation but a permanent ethical function. Participants' vigilance in verifying tone, inclusivity, and representational accuracy reflects the necessity of human judgment in maintaining equity and trust. "I always checked—does this sound like a nurse talking to a patient, or like a robot?" (P6).

Following Nguyen *et al.* (2023) and Bibi (2024), the ethical integration of AI must ensure that accountability resides with human actors empowered to intervene. Institutions should formalize these human verification roles as recognized positions—such as quality controllers, AI auditors, or instructional ethicists—rather than treating them as invisible labor. The LJF model demonstrates that when oversight is codified structurally, ethical transparency scales in tandem with the increasing use of automation.

5.3 Workforce Redefinition and Professional Identity

The distributed production model challenges traditional conceptions of authorship and ownership in instructional design. Participants expressed pride and cohesion—"Even though we each did a small part, it felt like one voice" (P8)—contradicting the assumption that automation fragments creativity. This shift aligns with the findings of Francis *et al.* (2024) and Cotilla Conceição and van der Stappen (2025), who link inclusion and collaboration to increased engagement.

Policy makers and academic leaders should recognize that emerging AI workflows require new competency frameworks: fluency in prompting, interpretive editing, and system feedback cycles. These hybrid roles—designers, reviewers, and AI collaborators—represent the future of instructional labor. Training and credentialing standards must adapt accordingly, treating human–AI partnership as a design specialization rather than an anomaly.

5.4 Data Integrity and Traceable Design

Every asset in the LJF workflow could be traced from objective to output, establishing an audit trail that satisfies emerging AI governance concerns around explainability. This operational transparency parallels the ARCHED model for responsible AI proposed by Li *et al.* (2025) and anticipates regulatory expectations for documentation in educational technology.

Policies should require that generative outputs in accredited learning environments maintain *objective provenance*: clear records linking each AI-generated element to its learning goal, authoring context, and human validation. Such traceability transforms compliance from burden to design feature—ensuring that quality assurance is embedded rather than external.

5.5 Institutional Scalability and Sustainability

The integration of multi-agent AI under the pedagogical discipline presents a replicable model for sustainable innovation. Across nineteen courses, the system consistently produced coherent, high-quality modules at unprecedented speed while maintaining alignment and rigor. This pattern parallels McCormick (2025), who emphasizes concept-based curricular reform as the pathway to sustainable nursing education. For institutions balancing financial pressure and pedagogical integrity, the findings demonstrate that efficiency and rigor do not have to conflict; when instructional frameworks structure automation, scalability becomes a function of alignment, not a threat to it.

5.6 Strategic Recommendations

To translate these insights into institutional and policy action:

- Mandate pedagogical frameworks—such as the LJF or comparable objective-driven architectures—for any large-scale AI course production.
- Codify human verification as a standing design role, ensuring ethical and representational review across all outputs.
- Require auditability by linking each AI-generated artifact to learning objectives and revision histories.
- Develop interdisciplinary training programs that pair instructional designers with AI technologists to maintain reciprocal fluency.
- Incentivize adaptive iteration, recognizing that continuous tuning—not static compliance—is the marker of sustainable quality.

The policy and practice implications converge on a central principle: AI can only scale education responsibly when governed by pedagogy rather than by code. The LJF operationalizes this philosophy, demonstrating that structured human oversight, objective mapping, and traceable workflows can yield efficiency without ethical compromise. As generative systems expand within education, these findings suggest that institutions should shift their policy attention from permission to use AI toward conditions under which AI is used to produce *learning responsibly.*

5.7 Significance of the Study

This study extends prior research on the LJF (Robinson, 2024, 2025a, 2025b) by empirically validating its scalability and ethical applicability within AI-assisted instructional design. The LJF's structured architecture became the foundation that enabled the AI system and human collaborators to produce rapid yet pedagogically coherent outcomes. The findings contribute to the growing body of evidence supporting pedagogically governed automation, demonstrating that when instructional logic disciplines AI, scalability, and rigor can coexist. The study holds significance for instructional designers, nursing program leaders, AI ethics researchers, and higher education administrators seeking sustainable frameworks for responsible innovation. By documenting the operation of a HITL system within a structured learning architecture, this research fulfills Creswell's call for integrative mixed-methods approaches that unify quantitative pattern and qualitative meaning (Creswell & Creswell, 2022).

5.8 Limitations and Future Research

While the findings affirm that an AI + HITL workflow can achieve both scalability and rigor, several boundaries frame the interpretation of these results. First, the study represents a single-institution pilot conducted within one disciplinary context—nursing education—and with a limited cohort of eight instructional contractors. As such, the results illustrate process validity rather than broad generalizability. The perspectives analyzed reflect those of professional practitioners embedded in the workflow rather than end users (students or faculty), and future studies should extend evaluation to learner outcomes and instructional delivery.

A second limitation concerns the proprietary nature of the multi-agent system. Although its architecture was pedagogically grounded in the LJF, the framework's creator continuously tuned its specific prompt logic, agent configurations, and automated QA cycles during production. This embedded authorship ensures fidelity but constrains replicability outside of contexts where the LJF is fully implemented. Replication studies should therefore examine whether similar outcomes can be achieved when the framework is applied independently of the system's original architect.

Third, measures of quality and efficiency relied primarily on perceptual and documentary evidence, including Likert-scale ratings, qualitative reflections, and revision logs. Although correction rates averaged below 5% of total assets, accuracy verification occurred through expert review rather than formal psychometric validation. Future research could incorporate objective performance metrics such as interrater reliability of alignment coding, Subject Matter Expert (SME) blind review scores, and downstream student indicators like National Council Licensure Examination (NCLEX) readiness or retention to triangulate perceptual and empirical quality measures.

Fourth, while approximately 30 specialized AI agents were orchestrated within this system, their individual contributions were evaluated collectively. Future analyses could disaggregate how specific agent types (e.g., those producing assessments versus media assets) influence perceived accuracy and workload. Similarly, comparative studies across disciplines may reveal whether domain knowledge density moderates the balance between automation and human revision.

Finally, the study did not examine long-term institutional adoption, cost–benefit tradeoffs, patterns of faculty uptake, or downstream student success, even though these factors are central to sustainability. Future research that includes cross-institutional pilots and faculty-led implementations should also track student performance indicators to determine whether the model improves learning outcomes over time. Replication in additional disciplines beyond nursing would further test the durability and transferability of this hybrid approach under varied curricular, accreditation, and workforce conditions.

In summary, the study provides early empirical grounding for how a pedagogically trained multi-agent AI can produce consistent, high-quality outputs under human oversight. Future research should shift from documenting feasibility to quantifying impact—connecting structural fidelity, accuracy rates, and learner performance outcomes—to establish the LJF not only as a design philosophy but also as an evidence-based standard for AI-enabled instructional development.

6 Conclusion

This study demonstrates that when AI is disciplined by pedagogical structure, scalability and rigor cease to be opposing forces. The LJF provided both the conceptual architecture and ethical boundaries through which a multi-agent, HITL system could operate at scale without eroding instructional integrity. Across nineteen courses and roughly 1,200 assets per course, participants described a process that was fast, structured, and human-centered—producing coherent, aligned, and pedagogically sound courses.

Quantitative results confirmed high consensus in efficiency, clarity of alignment, and perceived rigor. Meanwhile, qualitative evidence revealed that dialogic collaboration, shared prompting logic, and continuous system tuning sustained quality throughout rapid production cycles. Together, these strands illustrate that automation can be an extension of pedagogy rather than its dilution when governed by instructional intent. The AI did not design; it reasoned within the framework's boundaries, translating cognitive scaffolds into reproducible instructional patterns. Human reviewers remain essential not to repair errors, but to preserve voice, ethics, and representation, transforming oversight into a creative act of stewardship.

These findings extend the theoretical commitments advanced in prior work on the LJF (Robinson, 2025b) by demonstrating its operability as both a learning theory and a production system. The multi-agent model demonstrated that when each AI component is trained from a pedagogical standpoint, rather than merely a technical one, distributed intelligence can replicate the coherence of human design while expanding its reach. This aligns with Li *et al.* (2025), Memarian and Doleck (2024), and McCormick (2025), who collectively argue that sustainable innovation in education depends on frameworks that tether automation to instructional purpose.

Ultimately, the study reframes the discourse on AI in education. The question is no longer whether generative systems can produce learning materials, but under what structural and ethical conditions they do so responsibly. The LJF's success suggests that the future of AI in education will belong not to those who automate faster, but to those who design frameworks capable of teaching machines how to learn from pedagogy itself. In that synthesis—where human judgment, ethical transparency, and scalable design intersect—the promise of AI becomes not acceleration for its own sake, but amplification of what education has always sought to achieve: clarity, coherence, and meaningful learning at a human scale.

References

Allen, A. (2022). An Introduction to Constructivism: Its Theoretical Roots and Impact on Contemporary Education. *Journal of Learning Design and Leadership*, 1(1), 1-13.

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Addison Wesley Longman, Inc.

Bibi, A. (2024). Navigating The Ethical Landscape: AI Integration In Education. *Educational Administration Theory and Practice*, 30, 1579-1585. https://doi.org/10.53555/kuey.v30i6.5546

Braun, V., & Clarke, V. (2008). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Bruner, J. S. (1960). *The process of education*. Cambridge, Massachusetts: Harvard University Press.

Cotilla Conceição, J. M., & van der Stappen, E. (2025). The Impact of AI on Inclusivity in Higher Education: A Rapid Review. *Education Sciences*, 15(9), 1255. https://doi.org/10.3390/educsci15091255

Creswell, J. W., & Creswell, J. D. (2022). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Los Angeles: SAGE Publications.

Dazeley, R., Goriss-Hunter, A., Meredith, G., Sellings, P., Firmin, S., Burke, J., et al. (2024). Agile Backward Design: A Framework for planning higher education curriculum. *The Australian Educational Researcher*, 52(2), 1489-1508. https://doi.org/10.1007/s13384-024-00772-7

Edirisingha, P. (2024). Working with Generative AI to support interactive learning and teaching: reflections from the European Distance and E-Learning Conference in 2024. *Interactive Learning Environments*, 32(9), 4763-4765. https://doi.org/10.1080/10494820.2024.2414152

Francis, N. J., Jones, S., & Smith, D. P. (2024). Generative AI in Higher Education: Balancing Innovation and Integrity. *Br J Biomed Sci*, 81, 14048. https://doi.org/10.3389/bjbs.2024.14048

Han, S., Kang, H. S., Gimber, P., & Lim, S. (2025). Nursing Students' Perceptions and Use of Generative Artificial Intelligence in Nursing Education. *Nurs Rep*, 15(2), 68. https://doi.org/10.3390/nursrep15020068

Li, H., Fang, Y., Zhang, S., Lee, S. M., Wang, Y., Trexler, M., et al. (2025). ARCHED: A Human-Centered Framework for Transparent, Responsible, and Collaborative AI Assisted Instructional Design. *Proceedings of Machine Learning Research*, 273, 1-11. https://doi.org/10.48550/arXiv.2503.08931

Lungu, I. (2025). Backward Design – An Innovative Instructional Model in Planning Higher Education Courses. *Bulletin of the Transilvania University of Braşov. Series VII: Social Sciences • Law*, 17(66), 99-108. https://doi.org/10.31926/but.ssl.2024.17.66.4.9

McCormick, M. (2025). Strategic Curriculum Overhaul in a Bachelor of Science in Nursing Program: Integrating Concept-Based Frameworks, Competency-Based Assessment, and AI Innovation. *Medical Research Archives*, 13(9), 1-6. https://doi.org/10.18103/mra.v13i9.6920

Memarian, B., & Doleck, T. (2024). Human-in-the-loop in artificial intelligence in education: A review and entity-relationship (ER) analysis. *Computers in Human Behavior: Artificial Humans*, 2(1), 100053. https://doi.org/10.1016/j.chbah.2024.100053

Mosqueira-Rey, E., Hernández-Pereira, E., Alonso-Ríos, D., Bobes-Bascarán, J., & Fernández-Leal, Á. (2022). Human-in-the-loop machine learning: a state of the art. *Artificial Intelligence Review*, 56(4), 3005-3054. https://doi.org/10.1007/s10462-022-10246-w

Natarajan, S., Mathur, S., Sidheekh, S., Stammer, W., & Kersting, K. (2025). Human-in-the-loop or AI-in-the-loop? Automate or collaborate? In *Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence (AAAI-25)* (pp. 28594–28600). Association for the Advancement of Artificial Intelligence. https://doi.org/10.1609/aaai.v39i27.35083

Nguyen, A., Ngo, H. N., Hong, Y., Dang, B., & Nguyen, B. T. (2023). Ethical principles for artificial intelligence in education. *Educ Inf Technol (Dordr)*, 28(4), 4221-4241. https://doi.org/10.1007/s10639-022-11316-w

Retzlaff, C. O., Das, S., Wayllace, C., Mousavi, P., Afshari, M., Yang, T., et al. (2024). Human-in-the-Loop Reinforcement Learning: A Survey and Position on Requirements, Challenges, and Opportunities. *Journal of Artificial Intelligence Research*, 79, 359-415. https://doi.org/10.1613/jair.1.15348

Robinson, B. (2024). *Harnessing AI for Structured Learning: The Case for Objective-Driven Design in E-Learning*. International Journal on E-Learning 2024, Waynesville, NC, USA. https://www.learntechlib.org/p/225147

Robinson, B. (2025a). Implementing the Learner Journey Framework in a Foundational Writing Course: A Case Study in Scaffolded Design, AI Support, and Faculty Delivery. *Journal of Education and Practice*, 16(7), 27-38. https://doi.org/10.7176/jep/16-7-04

Robinson, B. (2025b). The Learner Journey Framework: A Constructivist Approach to Retention, Course Design, and Financial Sustainability in Higher Education. *Journal of Education and Practice*, 16(1), 138-148. https://doi.org/10.7176/jep/16-1-13

Roubides, P., & Roubides, A. (2025). Constructivism in Online Course Design of Skilled-based Courses and Its Impact on Student Achievement. *Journal of Research in Mathematics, Science, and Technology Education*, 2(1), 53-59. https://doi.org/10.70232/jrmste.v2i1.34

Schleiss, J., Laupichler, M. C., Raupach, T., & Stober, S. (2023). AI Course Design Planning Framework: Developing Domain-Specific AI Education Courses. *Education Sciences*, 13(9), 954. https://doi.org/10.3390/educsci13090954

Vygotsky, L. S. (1978). *Mind in Society: Development of Higher Psychological Processes*. Cambridge, Massachusetts: Harvard University Press.

Wiggins, G. P., & McTighe, J. (2005). *Understanding by Design*. Alexandria, Virginia: Association for Supervision and Curriculum Development.

Yao, C., Liu, S., Chen, H., & Wang, Y. (2025). Instructional agents: Multi-agent LLM framework for course material generation. *arXiv preprint arXiv:2308.12276 v4*.

Yao, H., Xu, W., Turnau, J., Kellam, N., & Wei, H. (2025). Instructional agents: LLM agents on automated course material generation for teaching faculties. *arXiv preprint arXiv:2508.19611*. https://doi.org/10.48550/arXiv.2508.19611

Table 1 – Course Asset Architecture

Asset Type	Description	Average per Course (10 Modules)
Module Introductions	Orient learners to module-level objectives, relevance, and NCLEX alignment	10
Module Outros	Summative reflections reinforcing connections between objectives and assessments	10
Text Pages	Core instructional content aligned to module- and enabling-level objectives	35 – 40
Videos	Short concept explanations and scenario-based demonstrations	20 – 25
Text-Based Interactives	Embedded comprehension checks and reflective prompts within text pages	~105
Video Interactives	In-video knowledge checks and decision points reinforcing key ideas	~75
Interactive Case Studies	Scenario-driven learning activities emphasizing application and transfer	30 – 35
Simulation Questions	Contextualized practice items appearing in three modules per course	12
NCLEX Practice Questions	Formative multiple-choice items mapped to NCLEX cognitive domains	10
Walk-With-Me Coaching Activities	Guided reasoning sequences aligned to NCLEX objectives	10
Discussion Prompts	Structured opportunities for reflection and peer engagement	10
Pulse Checks	Formative assessments—≈ 5 items per objective × 3 – 5 objectives per module	150 – 250
Images (Text Pages)	Instructional and conceptual visuals integrated with textual explanations	~120
Images (Videos)	Scenario and concept visuals embedded in instructional videos	~160
Active Reading Guides	Structured note-taking and synthesis tools supporting each module	10

NCLEX = The National Council Licensure Examination

Estimated Total Assets per Course: ≈ 1,200 instructional units

Courses Developed: 19 (produced under the AI + Human-in-the-Loop workflow)